基于地形因子和随机森林的丘陵区农田土壤有效铁空间分布预测

杨其坡[1,3]；武伟[2,3]；刘洪斌[1,3]

（1. 西南大学资源环境学院　重庆　400716; 2. 西南大学计算机与信息科学学院　重庆　400715; 3. 重庆市数字农业重点实验室　重庆　400716）

摘　要：为了掌握丘陵地区农田土壤有效铁含量及其空间分布，本文以重庆市江津区永兴镇内同源成土母质的典型丘陵(2 km^2)为研究区，采集 309 个土壤样点，利用普通克里格(Ordinary Kriging, OK)、多元线性回归(Multiple Linear Regression, MLR)、随机森林(Random Forest,, RF)模型，结合高程、坡度、坡向、谷深、平面曲率、剖面曲率、汇聚指数、相对坡位指数、地形湿度指数等地形因子对土壤有效铁进行空间分布预测，并通过 85 个验证点评价、筛选预测模型。结果表明：(1)土壤有效铁与谷深、地形湿度指数存在极显著水平正相关关系，与坡度、平面曲率、剖面曲率、汇聚指数、相对坡位指数存在极显著水平负相关关系；(2)随机森林模型的预测精度明显高于多元线性回归和普通克里格插值，其平均绝对误差为 22.33 mg kg^{-1}、均方根误差为 27.98 mg kg^{-1}、决定系数为 0.76，是研究区土壤有效铁含量空间分布的最适预测模型；(3)地形湿度指数和坡度是影响该区域土壤有效铁含量空间分布的主要地形因子。可见，基于地形因子的随机森林预测模型可以较好地解释丘陵区农田土壤有效铁含量的空间变异，研究结果为丘陵区土壤中、微量元素含量及空间分布预测提供方法借鉴和理论依据。

关键词：地形因子；随机森林模型；土壤有效铁；空间分布预测

中图分类号: S158.9　**文献标志码**: A

Prediction of spatial distribution of soil available iron at a typical hilly farmland based on terrain attributes and random forest model

YANG Qipo[1,3], WU Wei[2,3], LIU Hongbin[1,3**]

(1. College of Resources and Environment, Southwest University, Chongqing 400716, China; 2. College of Computer and Information Science, Southwest University, Chongqing 400715, China; 3. Chongqing Key Laboratory of Digital Agriculture, Chongqing 400716, China)

Abstract: Soil available iron is essential to plant growth**.** The detailed information about the spatial distribution of soil available iron will be important for effective management of soil fertility. To date, the published works mainly focused on investigating the spatial variability of soil available iron. Fewer of them were conducted to predict the spatial distribution of soil available iron. In order to understand the spatial distribution of soil available iron in hilly areas of southwest China, we conducted a study in 2014 at a typical hilly region which has uniform soil parent material and covers an area of 2 km^2 in Yongxing town, Jiangjin, Chongqing. A total of 309 samples were collected at a depth of 20 cm of the cultivated soils. They were randomly divided into calibration (224 samples) and validation (85 samples) datasets. Nine terrain attributes, namely, elevation, slope, aspect, valley depth, horizontal curvature, profile curvature, convergence index, relation position index, and topographic wetness index, were extracted from a digital elevation model with a spatial resolution of 2 m. The models of Ordinary Kriging (OK), Multiple Linear Regression (MLR), and Random Forest (RF) were used to predict the content of soil available iron based on the terrain attributes. Accuracy indicators including mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) were applied to evaluate model's performance based on the validation dataset. The results of correlation showed that the topographic wetness index and the valley depth had significant positive correlations with soil available iron. The slope, the horizontal curvature, the profile curvature, the convergence index and the relation position index had significant negative correlations with soil available iron. Compared with OK and MLR methods, RF model performed best with MAE of 22.33 mg kg^{-1}, RMSE of 27.98 mg kg^{-1}, and R^2 of 0.76. Additionally, the results of RF model indicated that the topographic wetness and the slope were the major factors controlling the spatial distribution of soil available iron. Therefore, Random Forest (RF) model combined with terrain attributes could explain the spatial variability of soil available iron over this area. The outcomes of the current work will provide valuable information for predicting the spatial distribution of the soil trace elements in hilly areas.

Keywords: terrain attribute; random forest model; soil available iron; prediction of spatial distribution

铁是植物体内细胞色素和酶的重要组成部分，在植物光合、呼吸等代谢过程中发挥着重要作用，进而影响作物的产量与品质[1-3]。土壤有效铁作为植物的可给态铁，是土壤对植物供铁能力的重要参考依据[4]。土壤有效铁的含量及其空间分布受到成土母质、气候、地形、土壤养分和土壤耕作方式等很多因素的影响[5-7]，造成了区域土壤有效铁含量变化幅度较大，影响植物的正常生长。丘陵是一种分布比较广泛的地貌类型，精准预测丘陵区土壤有效铁含量空间分布，了解土壤有效铁含量空间变异特征及其与地形之间的关系，对土壤地力评价和农用地利用规划具有极大的指导意义。然而，当前对于丘陵区土壤属性的空间分布预测研究也多见与土壤有机质和大量元素[8-9]，针对土壤有效铁的相关研究主要集中于空间特征分析。例如，廖琴[10]等研究了农田耕层区土壤有效铁的空间变异特征及分布规律，杜鹃[11]等对耕层土壤有效铁的丰缺状况和空间变异特征进行了研究，刘永红[12]等综合评价了土壤有效铁的空间格局变异特征。

至今，土壤属性空间预测的方法已经由传统土壤制图发展到数字化土壤制图，主要包括：地统计方法[13]、线性回归模型[14-15]和机器语言算法[16-18]。地统计学方法[19]在土壤属性空间预测中是以区域化变量为基础，空间相关和变异函数为工具的数学地质方法[20]。但是有时实际情况不符合二阶平稳或内蕴假设条件，造成不能可靠地应用地统计学来研究土壤属性空间预测的问题[21]；构建线性回归模型的前提条件是土壤属性与预测变量之间的关系是线性的，然而实际情况中两者的关系往往是非线性且复杂的[22]；机器语言算法：例如，决策树(Decision Trees, DT)[23]、支持向量机(Support Vector Machine, SVM)[24]、BP 神经网络(Back Propagation Neural Networks, BPNN)[25]等，近年来也被应用于土壤属性制图。然而这些模型易出现过度拟合[26-27]等缺陷。随机森林模型也属于机器语言算法的一种，它的出现在一定程度上克服了上述缺陷，提高了预测精度，是对土壤属性空间预测方法的有效改进和及时补充。郭澎涛[28]等利用随机森林模型结合环境变量对土壤全氮进行空间预测，其预测精度显著高于逐步线性回归模型、广义加性混合模型以及分类回归树模型。王茵茵[29]等利用遥感数据提取辅助环境因子结合随机森林算法对丘陵区表层土壤有机质进行空间预测，结果显示随机森林模型在复杂地貌区的预测更有效。

在一定区域尺度内，气候、成土母质等结构因素相对一致，地形对水热条件起到重分配作用并影响土壤发育[30]，是引起土壤属性空间变异的主要影响因素[31-32]。因此，近年来，很多关于土壤属性预测的研究都会考虑地形因子作为重要预测变量。连纲[33]等应用地形因子和遥感指数，分析土壤属性与环境的关系并进行空间预测；Grimm[34]等以地形因子和成土母质等环境因子为辅助变量对土壤有机碳进行了预测；张素梅[35]等利用地形因子和植被指数建立了土壤养分空间分布预测模型。

本研究旨在仅以地形因子为预测变量，利用随机森林模型对丘陵区土壤有效铁含量进行空间分布预测，为丘陵区土壤中、微量元素的空间分布预测提供方法借鉴和理论依据。

1　材料与方法

1.1　研究区概况

本文以重庆市江津区永兴镇内一个典型丘陵地带为研究区，经纬度区间为 106°09′27″~106°10′9″E，29°00′10″~29°00′9″N，面积约为2 km^2。研究区属于亚热带季风气候区，年平均气温18 ℃，年日照时数1253小时，年降水量900 mm，无霜期335 d。研究区为丘陵地貌，海拔238~328 m。研究区成土母质为侏罗系沙溪庙组紫色砂页岩，土地利用类型70%以上为耕地，主要是旱地和水田。按照土壤发生分类，旱地的土壤类型是紫色土类、中性紫色土亚类、灰棕紫泥土属；水田土壤类型是水稻土类、潜育水稻土亚类、潜育紫泥田土属。

1.2　土壤样品采集与化学分析

结合地形图和 1∶1000 土地利用现状图，依据均匀性、代表性原则，按照全国耕地地力调查与质量评价技术规程在研究区农田进行布点，在不同地形部位和坡向均要布设采样点，土壤样点共 309 个(旱地 258 个，水田 51 个)，样点分布如图 1 所示。2014 年 11 月进行土样采集，并记录土壤采样点坐标等信息，每个样点在同一样地中按照“S”形线路采样，取 10 钻(0~20 cm)耕作层土样用四分法混匀成约 1 kg 样品。将样品带回实验室风干、研磨、过筛，采用 DTPA 溶液浸提——原子吸收光谱法测定样品有效铁的含量[36]。

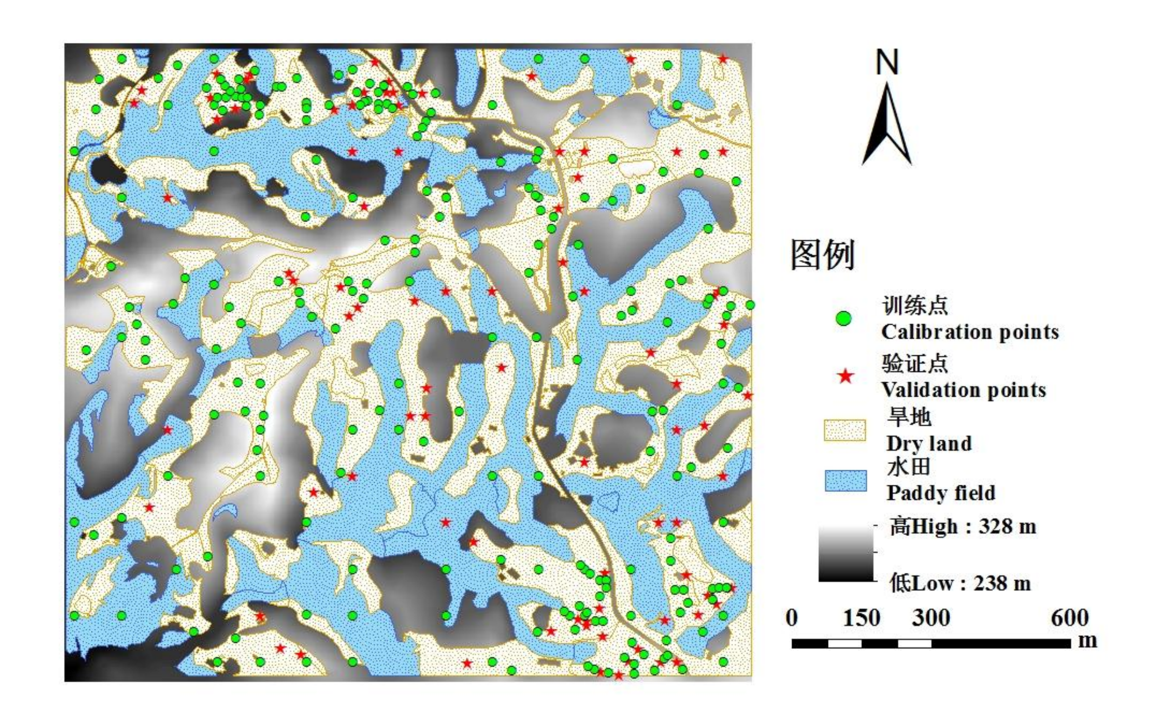


图 1 研究区 DEM、土地利用及样点分布图

Fig. 1 Maps of DEM, land use and samples sites of the study area

1.3 地形因子的提取

根据研究区尺度、地貌等信息，实地测量获得研究区 6373 个高程点和等高距为 1 m 的地形图。在 ArcGIS 中，使用 3D Analyst 模块，以高程点、等高线为输入变量创建 TIN 表面，依据 TIN 表面生成 2 m 分辨率的数字高程模型(Digital Elevation Model, DEM)。借鉴地形与土壤养分的相关研究[37]，本文选择 9 个地形因子：高程(Elevation, Ele)、坡度(Slope, Slo)、坡向(Aspect, Asp)、谷深(Valley depth, Vd)、平面曲率(Horizontal curvature, Hc)、剖面曲率(Profile curvature, Pc)、汇聚指数(Convergence index, Ci)、相对坡位指数(Relation position index, Rpi)、地形湿度指数(Topographic wetness index, Twi)。各地形因子含义见[38]，运用 SimDTA 软件[39]提取地形因子。参考 DEM 空间分辨率与地形信息之间关系[40-41]的相关研究，对原始 DEM 进行重采样，对比不同 DEM 分辨率(2 m、5 m、10 m、15 m、30 m)提取的地形因子对预测结果的影响，当 DEM 空间分辨率为 2 m 时，模型的预测精度最高，所以本文选择 2 m 空间分辨率的 DEM。

1.4 预测模型

1.4.1 普通克里格

普通克里格(Ordinary Kriging, OK)是地统计学的主要内容之一，以变异函数理论及结构分析为基础，根据已知采样点数据的线性组合，在一定区域内进行区域化变量取值的无偏最优估计[42]。

$$Z^*(x_0)=\sum_{i=1}^{n}\lambda_i\ Z(x_i)$$

式中: $Z^*(x_0)$表示待估点 x_0 的值, $Z(xi)$表示第 i 个有效观测值(i = 1,2,3…,n), λ_i 是通过半变异函数生成的权重且$\sum\lambda_i$=1。

1.4.2 多元线性回归

多元线性回归(Multiple Linear Regression, MLR)是基于普通最小二乘法，以多个解释变量的给定数据值为条件，研究一个因变量与多个自变量之间的线性关系[43]，普遍应用于土壤养分空间分布预测，其表达式为:

$$y_i=\beta_0+\beta_1x_{1i}+\beta_2x_{2i}+\beta_3x_{3i}\ldots+\beta_nx_{ni}$$

式中: x_i (i = 1,2,3,…,n)表示自变量, y_i (i = 0,1,2,…n)和 β_j (j = 0,1,2,…n)分别表示因变量和自变量的回归系数。

1.4.3 随机森林模型

随机森林(Random Forest, RF)是一种基于分类树的多功能机器语言学习算法, 运用 bootstrap 重抽样法从原始样本中随机抽取多个样本进行决策树模型的构建, 通过对多棵决策树的预测进行投票得到最终结果。随机森林回归时对数据分布、数据类型、数据结构没有要求, 并对噪声和异常值有良好的容忍度, 并且不容易有过度拟合的状况[44]。

1.5 半方差函数

半方差函数是研究土壤属性变异性的函数, 可以确定土壤属性的空间相关性[45], 其表达式为:

$$r(h) = \frac{1}{2N(h)}\sum_{i=1}^{N(h)}[z(x_i + h) - z(x_i)]^2$$

式中: $r(h)$表示间距为 h 的点对之间的平均半方差, $N(h)$表示距离为 h 时的点对数量, $z(x_i)$则表示点 x_i 的观测值。

1.6 预测模型构建及精度评价

本研究随机选择 224 个采样点(旱地 190 个, 水田 34 个)作为训练集用于预测模型的构建, 余下 85 个样点(旱地 68 个, 水田 17 个)作为验证集用来评价模型的预测精度。普通克里格方法根据已知点数据估计待测点, 不需要加入预测变量。根据半方差分析得到的指数模型, 在 ArcGIS 软件中进行插值。多元线性回归通过 SPSS 软件计算, 随机森林模型使用 R 软件中 RandomForest 包[46]进行模拟, 随机森林中参数 ntree(决策树的数量)和 mtry(每次树重建时节点分裂的次数)最终设定为分别为 150、3。

模型预测精度的评价指标包括: 平均绝对误差(mean absolute error, MAE)、均方根误差(root mean square error, RMSE)和决定系数(coefficient of determination, R^2)。MAE、RMSE 越小, R^2 越大, 表明模型预测精度越高。相关计算公式如下:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|M_i - P_i| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(M_i - P_i)^2} \quad (2)$$

$$R^2 = \sum_{i=1}^{n}(P_i - \bar{M})^2 \,/\, \sum_{i=1}^{n}(M_i - \bar{M})^2 \quad (3)$$

式中: M 为实测值, P 为预测值, $\bar{M}$为实测值均值, n 则为验证集样本数量。

2 结果与分析

2.1 土壤有效铁描述性统计

研究区土壤有效铁含量统计结果如表 1 所呈现。训练集土壤有效铁的含量范围(3.00~276.97 mg kg^{-1})稍大于验证集有效铁的含量范围(4.20~208.38 mg kg^{-1}), 训练集土壤有效铁含量的标准差(57.35 mg kg^{-1})偏高于验证集有效铁含量的标准差(53.85 mg kg^{-1})。方差分析结果表明训练集与验证集土壤有效铁含量差异不显著, 表明训练数据和验证数据对研究区土壤有效铁含量都具有较好的代表性。训练集和验证集的土壤有效铁变异系数均大于 1, 说明各采样点之间土壤有效铁含量存在明显差异, 属于强变异性, 与徐小逊等[47]采用地统计学方法研究沱江中游耕层土壤有效铁的空间分布时结果类似。训练集和验证集土壤有效铁含量的偏度都大于 1, K-S 检验结果 p 值小于 0.05, 表明研究区土壤有效铁含量不服从正态分布, 对数转换后的土壤有效铁服从正态分布, 可用于多重线性回归模拟和普通克里格插值。

表 1 研究区采样点土壤有效铁含量描述性统计

Table 1 Descriptive statistics of soil available iron

样本组 Sample group	样本数 Number of samples	最小值 Minimum (mg kg^{-1})	最大值 Maximum (mg kg^{-1})	中值 Median (mg kg^{-1})	平均值 Mean (mg kg^{-1})	标准差 Std.Deviation (mg kg^{-1})	变异系数 Coefficient of variation	偏度 Skewness

训练集 Calibration samples	224	3.00	276.97	27.36	49.94	57.35	1.15	1.80
验证集 Validation Samples	85	4.20	208.38	26.44	49.55	53.85	1.09	1.49

2.2　土壤有效铁的半方差分析

本研究中根据土壤有效铁含量的统计结果，应用 GS+软件进行半方差计算，得到半方差最优拟合模型及特征参数见表 2。表中块基比即基底效应，表示数据的空间相关性强弱。该值小于 25%为较强的空间相关性，大于 75%为较弱的空间相关性，介于 25%-75%表示中等空间自相关。土壤有效铁的基底效应为 14.5%，说明具有较强的空间相关性，表明采样点之间土壤有效铁含量的差异主要受到结构性因素的影响，如气候、成土母质、地形等。因为该研究区属小尺度丘陵区域，成土母质相同、气候条件一致，所以表明地形是影响土壤有效铁空间分布的主要因素。

表 2　土壤有效铁的半方差模型参数

Table 2　Semivariance parameters of soil available iron

	拟合模型 Model	块金值 Nugget (C_0)	基台值 Sill (C_0+C)	块基比 Proportion [C_0/(C_0+C)] %	变程 Range (m)	决定系数 R^2
土壤有效铁 Soil available iron	指数模型 Exponential	560	3858	14.5	561	0.898

2.3　土壤有效铁含量与地形因子的相关性

表 3 为土壤有效铁与地形因子之间的相关性。土壤有效铁与坡度、谷深、平面曲率、剖面曲率、汇聚指数、相对坡位指数、地形湿度指数均达到极显著水平相关关系，表明在研究区土壤有效铁受到地形因素的影响明显。土壤有效铁与谷深、地形湿度指数存在极显著水平正相关关系，相关系数分别是 0.298、0.592，表明在地势低、土壤含水量大的区域有利于土壤有效铁的积累；土壤有效铁与坡度、平面曲率、剖面曲率、汇聚指数、相对坡位指数存在极显著水平负相关关系，相关系数分别是–0.371、–0.327、–0.228、–0.174、–0.428，表明在地势高、起伏大的区域，土壤有效铁的含量较低。

表 3　土壤有效铁与地形因子的相关性

Table 3　Correlations between soil available iron and terrain attributes

	Ele	Slo	Asp	Vd	Hc	Pc	Ci	Rpi	Twi
土壤有效铁 Soil available iron	–0.116	–0.371**	0.086	0.298**	–0.327**	–0.228**	–0.174**	–0.428**	0.592**

注：表中的数据为训练集统计分析结果。Ele 为高程；Slo 为坡度；Asp 为坡向；Vd 为谷深；Hc 为平面曲率；Pc 为剖面曲率；Ci 为汇聚指数；Rpi 为相对坡位指数；Twi 为地形湿度指数。*和**表明地形因子与土壤有效铁的相关性达到了显著($P<0.05$)和极显著($P<0.01$)水平。

Note: the results were calculated based on the data of training of set. Where Ele is the elevation, Slo is the slope, Asp is the aspect, Vd is the valley depth, Hc is the horizontal curvature, Pc is the profile curvature, Ci is the convergence index, Rpi is the relation position index, Twi is the Topographic wetness index. * and ** stand for that the relationship between terrain attributes and soil available iron reached the significant level ($P<0.05$) and highly significant level ($P<0.01$), respectively.

2.4　土壤有效铁含量在不同土地利用类型的对比

表 4 为土壤有效铁在研究区不同土地利用类型的对比。由表可知，旱地样本数(190 个)远远大于水田(34 个)，这是因为旱地与水田相比地形较复杂，地块小且破碎，旱地土壤有效铁含量的变化可能更易受地形的影响。经方差分析，土壤有效铁在水田中含量(151.04±11.54 mg kg^{-1})明显大于旱地(31.85±2.20 mg kg^{-1})，且差异达到了显著性水平。同时旱地土壤有效铁的变异系数(0.95)大于水田(0.45)，但都属于中等变异性(变

异系数在 0.1~1 之间为中等变异)，说明在同一土地类型中各采样点之间土壤有效铁含量存在一定差异。

表 4　土壤有效铁在旱地和水田之间的比较

Table 4　Comparison of soil available iron between cropland and paddy field

土地利用类型 Land use type	样本数 Number of samples	最小值 Minimum /(mg kg^{-1})	最大值 Maximum /(mg kg^{-1})	中值 Median /(mg kg^{-1})	平均值 Mean /(mg kg^{-1})	标准差 Std.Deviation /(mg kg^{-1})	变异系数 Coefficient of variation
旱地 Dry land	190	3.00	156.60	21.24	31.85	30.37	0.95
水田 Paddy field	34	3.85	276.97	155.22	151.04	67.27	0.45

注：表中的数据为训练集统计分析结果。

Note: the results were calculated based on the data of training set.

2.4　土壤有效铁含量模型预测精度比较

研究区土壤有效铁含量模型预测精度通过 85 个采样点(旱地 68 个，水田 17 个)进行验证，结果见表 5。随机森林(RF)模型的平均绝对误差 MAE(22.33 mg kg^{-1})、均方根误差 RMSE(27.98 mg kg^{-1})都明显低于普通克里格(OK)、多元线性回归(MLR)，决定系数 R^2(0.76)明显高于普通克里格(OK)、多元线性回归(MLR)。这是因为多元线性回归(MLR)模型只能拟合土壤有效铁与地形因子之间的基本线性关系，而对非线性关系难以预测，制约了模型预测精度的提高；普通克里格(OK)模型预测精度较低，原因在于插值方法仅利用了土壤有效铁的空间自相关性，不能顾全研究区地形等环境的影响。随机森林(RF)模型充分考虑了地形对土壤有效铁的影响，并且捕捉到了它们之间除线性之外的复杂关系，预测精度较高。研究区土壤有效铁含量模型预测精度结果进一步表明了随机森林(RF)模型自身的优越性和利用地形因子预测小区域尺度丘陵农田土壤有效铁含量空间分布的可行性。

由土壤有效铁在旱地和水田的比较可知，不同土地利用类型(旱地、水田)的土壤有效铁含量差异显著，所以构建预测模型时需要考虑土地利用类型的影响。因为土地利用类型属于定性变量，随机森林模型相比多元线性回归模型对定性变量具有更好的包容性，可以较多地捕捉土地利用类型与土壤有效铁之间的有效信息。同时通过表 5 的土壤有效铁含量预测模型精度比较可知，随机森林模型相比多元线性回归模型展现出较好的预测能力。于是挑选出具有较好预测精度的随机森林模型，在其预测变量中加入土地利用类型(旱地、水田)重新构建模型，检验土地利用类型是否可以进一步提高随机森林模型的预测精度。基于地形和土地利用的随机森林预测模型的平均绝对误差(MAE)为 23.77 mg kg^{-1}、均方根误差(RMSE)为 32.95 mg kg^{-1}、决定系数(R^2)为 0.66。结果显示随着预测变量中土地利用类型(旱地、水田)的加入，随机森林模型预测精度下降。这是因为水田、旱地的分布与地形密切相关[48-50]。土地利用类型是地形的一种综合、整体、定性表现形式，水田、旱地对其他地形因子造成了干扰、遮盖及共线性影响，使得各个地形因子对模型的贡献无法全部体现，导致模型预测误差增大，从而降低了模型的预测精度。综合模型精度评价指标，该研究区土壤有效铁含量的最适模型是基于地形因子的随机森林模型。

表 5　土壤有效铁含量模型预测精度

Table 5　Prediction accuracy of fitted models of soil available iron

预测模型 Prediction model	验证集 Validation dataset MAE (mg kg^{-1})	RMSE (mg kg^{-1})	R^2
OK	32.97	52.38	0.11
MLR	23.58	35.25	0.60
RF	22.33	27.98	0.76

2.5　地形因子对土壤有效铁含量空间变异的影响

图 2 展示了研究区地形因子对随机森林土壤有效铁含量预测模型的影响。地形作为五大成土因素之一，

影响土壤与环境之间的物质、能量交换，引起土壤的理化性质和养分变化[51-52]。由图可知，影响研究区土壤有效铁含量空间分布的主要地形因子是地形湿度指数(Twi)和坡度(Slo)。坡度通过影响区域降雨及入渗时间、表层土壤颗粒运动、径流挟沙能力和侵蚀方式等来间接影响土壤肥力特征[53-55]。地形湿度指数可定量模拟土壤水分含量状况和径流产生潜在能力，其值愈大，代表土壤含水量愈大[56-58]。由土壤有效铁含量与地形因子的相关性分析可知，土壤有效铁含量与地形湿度指数(Twi)呈极显著正相关(0.592)；与坡度(Slo)呈极显著负相关(-0.371)。表明随着坡度的增加，地表陡峭程度增加，降雨对土壤的侵蚀程度加重，通过淋溶作用造成土壤养分(土壤有效铁、全量铁、有机质等)的流失严重。土壤全量铁含量过低会直接导致有效铁含量的不足，同时土壤中有机质的缺乏也会引起土壤有效铁的缺失。随着地形湿度指数的增加，土壤含水量增加，导致土壤空气相对不足。在这种通气不良的条件下，土壤的还原性增强，Fe^{3+}易于转换成 Fe^{2+}，促进了土壤有效铁的增加[59-60]。

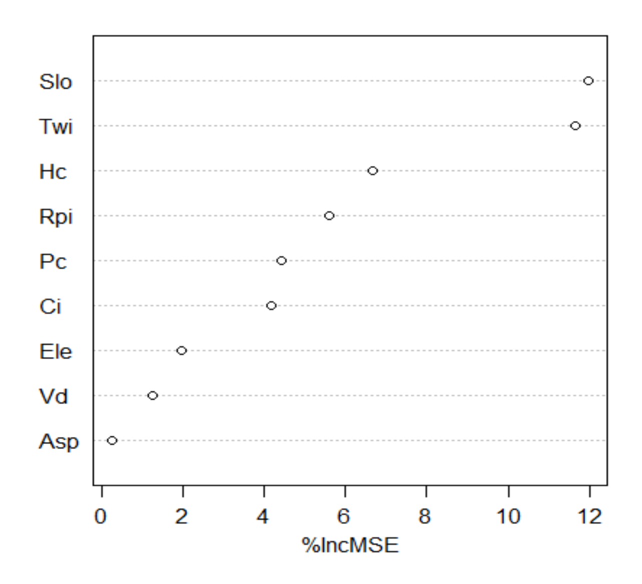


图 2 随机森林自变量影响力评价

Fig.2 Influence evaluation of independent variables by Random Forest

图中 IncMSE 为均方误差增加量; Ele 为高程; Slo 为坡度; Asp 为坡向; Vd 为谷深; Hc 为平面曲率; Pc 为剖面曲率; Ci 为汇聚指数; Rpi 为相对坡位指数; Twi 为地形湿度指数。In the graph, IncMSE is increase in mean squared error; Ele is the elevation; Slo is the slope; Asp is the aspect; Vd is the valley depth; Hc is the horizontal curvature; Pc is the profile curvature; Ci is the Convergence index; Rpi is the relation position index; Twi is the Topographic wetness index.

2.6 土壤有效铁空间分布预测

应用以地形因子为自变量的随机森林预测模型，在 ArcGIS 中得到土壤有效铁含量空间分布图(图 3)。由图可知，土壤有效铁含量分布趋势与研究区地形密切相关。地势陡峭的区域，土壤有效铁含量较低，这是由于降雨引起的土壤侵蚀、淋溶加重，造成土壤有效铁和其他土壤养分的流失，进而引起土壤有效铁含量的降低。地形低洼的区域，土壤有效铁含量较高，这是因为流失的土壤养分会在低洼处聚集，进一步积累。同时地形低洼处，土壤含水量大，造成了土壤的还原条件，更有利于土壤有效铁含量的增加。

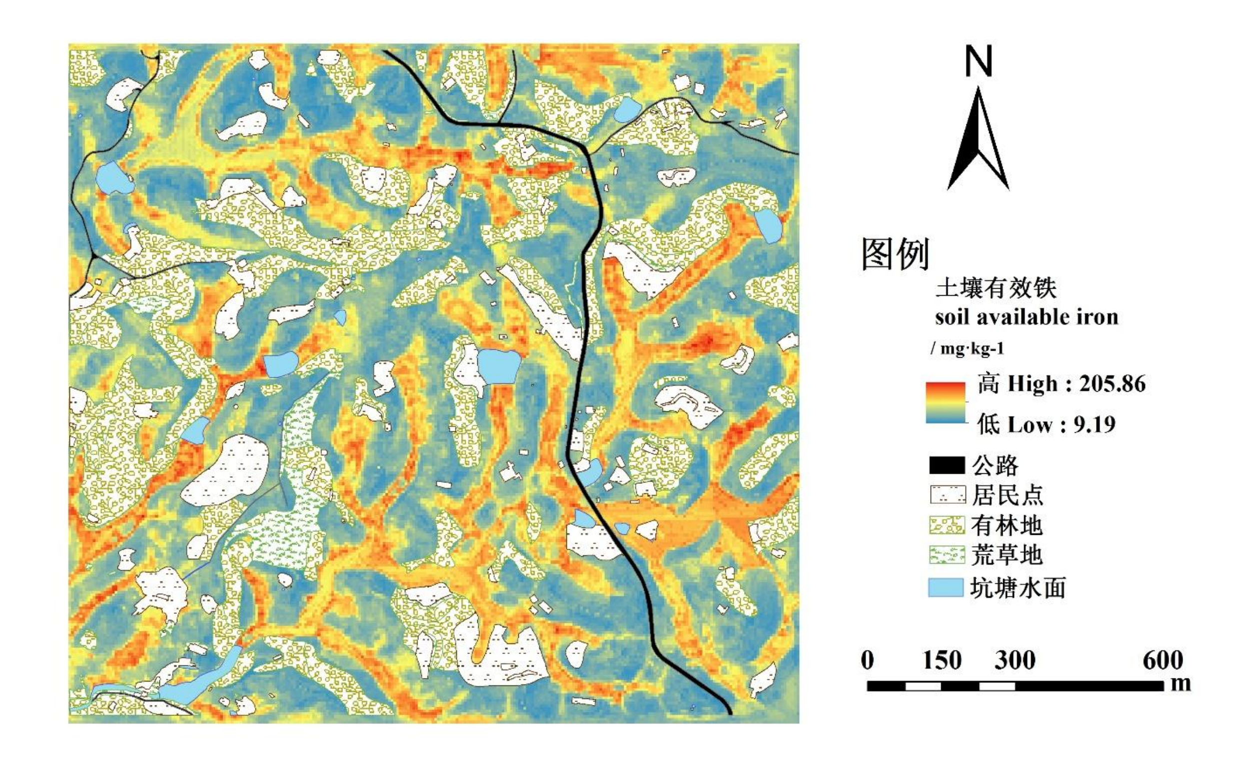


图 3 土壤有效铁含量空间分布图

Fig. 3 Prediction map of soil available iron

3 结论

本文首先以地形因子为预测变量，通过普通克里格、多元线性回归、随机森林等模型对研究区土壤有效铁的空间分布进行预测，筛选出结果较好的随机森林模型。然后通过对比基于地形-土地利用的随机森林模型，最终得出该研究区土壤有效铁含量的最适预测模型是基于地形因子的随机森林模型。

(1)研究区地形与土壤有效铁含量空间分布密切相关，土壤有效铁与坡度、谷深、平面曲率、剖面曲率、汇聚指数、相对坡位指数、地形湿度指数均达到极显著水平相关关系。土壤有效铁含量空间变异主要受到结构性因素的影响。

(2)利用以地形因子为自变量的随机森林模型，其平均绝对误差 MAE 为 22.33 mg kg^{-1}、均方根误差 RMSE 为 27.98 mg kg^{-1}、决定系数 R^2 为 0.76，显著优于其他预测模型，可以作为一种新的模型利用地形因子对土壤中、微量元素含量进行空间分布预测。

(3)地形湿度指数和坡度是影响该区域土壤有效铁含量空间分布的主要地形因子。

随机森林模型能较好捕捉该研究区地形因子和土壤有效铁含量之间的关系，可以利用地形因子解释土壤有效铁含量空间变异的 76%，但是预测精度有待进一步提高。在未来的研究中，应该考虑加入其它环境变量并筛选预测变量进行随机森林模型的建模，提高其在土壤中、微量元素空间分布的预测精度。

参考文献 References

[1] 汪李平. 植物的铁素营养及缺铁症的防治(综述)[J]. 安徽农业大学学报,1995,(01):17-22.

Wang L P. Review on iron nutrition and control of iron deficiency in plants[J]. Journal of Anhui Agricultural University, 1995,(01):17-22.

[2] 申红芸，熊宏春，郭笑彤，等. 植物吸收和转运铁的分子生理机制研究进展[J]. 植物营养与肥料学报，2011, 17(6):1522-1530.

Shen H Y, Xiong H C, Guo X T, et al. Progress of molecular and physiological mechanism of iron uptake and translocation in

plants[J]. Plant Nutrition & Fertilizer Science, 2011, 17(6):1522-1530.
[3] 李俊成，于慧，杨素欣,等. 植物对铁元素吸收的分子调控机制研究进展[J]. 植物生理学报, 2016(6):835-842.
Li J C, Yu H, Yang S X, et al. Research progress of molecular regulation of iron uptake in plants[J]. Plant Physiology Journal, 2016(6):835-842.
[4] 张春梅. 古浪灌区土壤耕层有效锌、锰、铜、铁含量分析与评价[J]. 草业科学,2011,(06):1221-1225.
Zhang C M. Analysis and evaluation of available Zn, Mn, Cu and Fe contents of topsoil in Gulang Irrigation Region[J]. Pratacultural Science, 2011, 28(6):1221-1225.
[5] 李强，周冀衡，杨荣生,等. 曲靖植烟土壤养分空间变异及土壤肥力适宜性评价[J]. 应用生态学报, 2011, 22(4):950-956.
Li Q, Zhou J H, Yang R S, et al. Soil nutrients spatial variability and soil fertility suitability in Qujing tobacco-planting area[J]. Chinese Journal of Applied Ecology, 2011, 22(4):950-956.
[6] 武婕，李玉环，李增兵,等. 南四湖区农田土壤有机质和微量元素空间分布特征及影响因素[J]. 生态学报，2014, 34(6):1596-1605.
Wu J, Li Y H, Li Z B, et al. Spatial distribution and influencing factors of farmland soil organic matter and trace elements in the Nansihu Region[J]. Acta Ecologica Sinica, 2014, 34(6):1596-1605.
[7] Jenny H. The soil resource. Origin and behavior[J]. Vegetatio, 1984, 57(2-3):102-102.
[8] 黄安，杨联安，杜挺,等. 基于多元成土因素的土壤有机质空间分布分析[J]. 干旱区地理(汉文版), 2015, 38(5):994-1003.
Huang A, Yang L A, Du T, et al. Spatial distribution of the soil organic matter based on multiple soil factors[J]. Arid Land Geography, 2015, 38(5):994-1003.
[9] 李启权,王昌全,张文江,余勇,李冰,杨娟,白根川,蔡艳. 基于神经网络模型和地统计学方法的土壤养分空间分布预测[J]. 应用生态学报,2013,(02):459-466.
Li Q Q, Wang C Q, Zhang W J, et al. [Prediction of soil nutrients spatial distribution based on neural network model combined with goestatistics][J]. Chinese Journal of Applied Ecology, 2013, 24(2):459-466.
[10] 廖琴,南忠仁,王胜利,黄璜,晋王强,周婷. 干旱区绿洲农田土壤微量元素有效态含量空间分布特征[J]. 环境科学研究,2011,(03):273-280.
Liao Q, Nan Z R, Wang S L, et al. Spatial distribution characteristics of available microelement contents in oasis cropland soils of arid areas[J]. Research of Environmental Sciences, 2011, 24(3):273-280.
[11] 杜娟,张永清,周进财,李明. 尧都区耕层土壤微量元素有效态含量空间分布特征[J]. 中国农学通报,2014,(03):162-167.
Du J, Zhang Y Q, Zhou J C, Li M, et al. Spatial distribution characteristics of available Microelements contents in irrigation-silted soil in Yaodu District[J]. Chinese Agricultural Science Bulletin, 2014, 30(3):162-167.
[12] 刘永红,倪中应,谢国雄,徐立军,钟林炳,马立强. 浙西北丘陵区农田土壤微量元素空间变异特征及影响因子[J]. 植物营养与肥料学报,2016,(06):1710-1718.
Liu Y H, Ni Z Y, Xie G X, Xu L J, Zhong L B, Ma L Q. Spatial variability and impacting factors of trace elements in hilly region of cropland in northwestern Zhejiang Province[J].Journal of Plant Nutrition and Fertilizer,2016,(06):1710-1718.
[13] 史文娇,岳天祥,石晓丽,宋伟. 土壤连续属性空间插值方法及其精度的研究进展[J]. 自然资源学报,2012,(01):163-175.
Shi W J, Yue T X, Shi X L, et al. Research progress in soil property interpolators and their accuracy[J]. Journal of Natural Resources, 2012,(01):163-175.
[14] 张国平，郭澎涛，王正银,等. 紫色土丘陵地区农田土壤养分空间分布预测[J]. 农业工程学报, 2013, 29(6):113-120.
Zhang G P, Guo P T, Wang Z Y, Liu H B,et al. Prediction of spatial distribution of hilly farmland with purple soil nutrient[J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(6):113-120.
[15] WANG J, CUI L, GAO W, et al. Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy[J]. Geoderma, 2014, 216(4): 1 - 9.
[16] 徐剑波，宋立生，夏振，等. 基于 GARBF 神经网络的耕地土壤有效磷空间变异分析[J]. 农业工程学报，2012, 28(16):158-165.
Xu J B, Song L S, Xia Z, et al. Spatial variability of available phosphorus for cultivated soil based on GARBF neural network[J]. Transactions of the Chinese Society of Agricultural Engineering, 2012, 28(16):158-165.
[17] 石伟,南卓铜,李韧,赵林,张秀敏,赵拥华. 基于支持向量机的典型冻土区土壤制图研究[J]. 土壤学报,2011,(03):461-469.
Shi W, Nan Z T, Li R, Zhan L, Zhang X M, Zhan Y H. Support vector machine based soil mapping of a typical permafrost area in the Qinghai-Tibet plateau[J]. Acta Pedologica Sinica, 2011,(03):461-469.
[18] 芦园园,张甘霖,赵玉国,等. 复杂景观环境下土壤厚度分布规则提取与制图[J]. 农业工程学报,2014,(18):132-141.
Lu Y Y, Zhang G L, Zhao Y G, et al. Extracting and mapping of soil depth distribution rules in complex landscape environment[J]. Transactions of the Chinese Society of Agricultural Engineering, 2014, 30(18):132-141.
[19] Lark R M. Towards soil geostatistics[J]. Spatial Statistics, 2012, 1:92-99.

[20] Timothy C. Coburn. Geostatistics for Natural Resources Evaluation[J]. Journal of Environmental Quality, 2012, 42(4):437-438.
[21] 徐剑波,宋立生,彭磊. 土壤养分空间估测方法研究综述[J]. 生态环境学报,2011,(Z2):1379-1386.
Xu J B, Song L S, Peng L. Research review on methods of spatial prediction of soil nutrients[J]. Ecology & Environmental Sciences, 2011, (Z2):1379-1386.
[22] 丁建丽，王飞. 干旱区大尺度土壤盐度信息环境建模——以新疆天山南北中低海拔冲积平原为例[J]. 地理学报，2017, 72(1):64-78.
Ding J L, Wang F. Environmental modeling of large-scale soil salinity information in an arid region:A case study of the low and middle altitude alluvial plain north and south of the Tianshan Mountains, Xinjiang[J]. Acta Geographica Sinica, 201772(1):64-78.
[23] Breiman L I, Friedman J H, Olshen R A, et al. Classification and Regression Trees (CART)[J]. Biometrics, 2015, 40(3):358.
[24] Besalatpour A, Hajabbasi M, Ayoubi S, et al. Prediction of soil physical properties by optimized support vector machines[J]. International Agrophysics, 2012, 26(2):109-115.
[25] Koley S. Machine Learning for Soil Fertility and Plant Nutrient Management Using Back Propagation Neural Networks[J]. Social Science Electronic Publishing, 2016, 2(2):292-297.
[26] Dietterich T. Overfitting and undercomputing in machine learning[J]. Acm Computing Surveys, 1995, 27(3):326-327.
[27] Hawkins D M. The problem of overfitting.[J]. Cheminform, 2004, 44(1):1.
[28]郭澎涛，李茂芬，罗微,等. 基于多源环境变量和随机森林的橡胶园土壤全氮含量预测[J]. 农业工程学报，2015, 31(5):194-202.
Guo P T, Li M F, Luo L, et al. Prediction of soil total nitrogen for rubber plantation at regional scale based on environmental variables and random forest approach[J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(5):194-202.
[29] 王茵茵,齐雁冰,陈洋,等. 基于多分辨率遥感数据与随机森林算法的土壤有机质预测研究[J]. 土壤学报,2016,(02):342-354.
Wang Y Y, Qi Y B, Chen Y, et al. Prediction of Soil Organic Matter Based on Multi-resolution Remote Sensing Data and Random Forest Algorithm[J]. Acta Pedologica Sinica, 2016, (02):342-354.
[30] 黄昌勇. 土壤学[M]. 北京:中国农业出版社,1999,141-142.
Huang C Y. Soil Science[M]. Beijing: China Agricul Ture Press, 1999, 141-142.
[31] 秦松，樊燕，刘洪斌,等. 地形因子与土壤养分空间分布的相关性研究[J]. 水土保持研究, 2008, 15(1):275-279.
Qin S, Fan Y, Liu H B, et al. Study on the Relations Between Topographical Factors and the Spatial Distributions of Soil Nutrients[J]. Research of Soil & Water Conservation, 2008, 15(1):275-279.
[32] 宋轩，李立东，寇长林,等. 黄水河小流域土壤养分分布及其与地形的关系[J]. 应用生态学报, 2011, 22(12):3163-3168.
Song X, Li L D, Kou C L, et al. [Soil nutrient distribution and its relations with topography in Huangshui River drainage basin][J]. Chinese Journal of Applied Ecology, 2011, 22(12):3163-3168.
[33] 连纲，郭旭东，傅伯杰,等. 基于环境相关法和地统计学的土壤属性空间分布预测[J]. 农业工程学报, 2009, 25(7):237-242.
Lian G, Guo X D, Fu B J, et al. Prediction of the spatial distribution of soil properties based on environmental correlation and geostatistics.[J]. Transactions of the Chinese Society of Agricultural Engineering, 2009, 25(7):237-242.
[34] Grimm R, Behrens T, Märker M, et al. Soil organic carbon concentrations and stocks on Barro Colorado Island-Digital soil mapping using Random Forest analysis [J]. Geoderma, 2008, 146(1-2): 102-113.
[35] 张素梅,王宗明,张柏,等. 利用地形和遥感数据预测土壤养分空间分布[J]. 农业工程学报,2010,(05):188-194.
Zhang S M, Wang Z M, Zhang B, et al. Prediction of spatial distribution of soil nutrients using terrain attributes and remote sensing data.[J]. Transactions of the Chinese Society of Agricultural Engineering, 2010, 26(5):188-194.
[36] 鲍士旦. 土壤农化分析 [M]. 北京:中国农业出版社,2002,47－56.
Bao S D. Soil and Agricultural Chemistry Analysis[M]. Beijing: China Agricul Ture Press, 2002,47－56.
[37] 邓欧平,周稀,黄萍萍,等. 川中紫色丘区土壤养分空间分异与地形因子相关性研究[J]. 资源科学,2013,(12):2434-2443.
Deng O P, Zhou X, Huang P P, et al. Correlations between spatial variability of soil nutrients and topographic factors in the purple hilly region of Sichuan[J]. Resources Science, 2013, (12):2434-2443.
[38] Moore I D, Gessler P E, Nieslen G A, et al. Soil attribute prediction using terrain analysis[J]. Soil Sci Soc Am J, 1993, 57: 443－452.
[39] 秦承志,卢岩君,包黎莉,朱阿兴,等. 简化数字地形分析软件(SimDTA)及其应用——以嫩江流域鹤山农场区的坡位模糊分类为例[J]. 地球信息科学学报,2009,(06):737-743.
Qin C Z, Lu Y J, Bao L L, Zhu A X, et al. Simple Digital Terrain Analysis Software (SimDTA 1.0) and Its Application in Fuzzy Classification of Slope Positions[J]. Journal of Geo-Information Science, 2009, 11(6):737-743.
[40] 呼雪梅，秦承志. 地形信息对确定 DEM 适宜分辨率的影响[J]. 地理科学进展, 2014, 33(1):50-56.

Hu X M, Qin C Z. Effects of different topographic attributes on determining appropriate DEM resolution[J]. Progress in Geography, 2014, 33(1):50-56.
[41] 黄骁力，汤国安，刘凯. DEM 分辨率对地形纹理特征提取的影响[J]. 地球信息科学学报, 2015, 17(7):822-829.
Huang X L, Tang G A, Liu K. The Influence of DEM resolution on the extraction of terrain texture feature[J]. 地球信息科学学报, 2015, 17(7):822-829.
[42] Dale L. Zimmerman, M. Bridget Zimmerman. A Comparison of Spatial Semivariogram Estimators and Corresponding Ordinary Kriging Predictors[J]. Technometrics, 2012, 33(1):77-91.
[43] Fedotova O, Teixeira L, Alvelos H. Software Effort Estimation with Multiple Linear Regression: review and practical application[J]. Journal of Information Science & Engineering, 2013, 29(5):925-945.
[44] Lindner C, Bromiley P A, Ionita M C, et al. Robust and Accurate Shape Model Matching Using Random Forest Regression-Voting[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9):1862-1874.
[45] Goovaerts P. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties[J]. Biology & Fertility of Soils, 1998, 27(4):315-334.
[46] Breiman L, Cutler A. Breiman and Cutler s random forests for classification and regression. R Package Version 4.6-7,2013.
[47] 徐小逊,张世熔,余妮娜,等. 沱江中游土壤有效铁空间分布及其影响因素分析[J]. 西南农业学报,2012,(03):977-981.
Xu X X, Zhang S R, Yu N N, et al. Soil available iron spatial distribution and influencing factors analysis based on GIS in middle reaches of Tuojiang[J]. Southwest China Journal of Agricultural Sciences, 2012,(03):977-981.
[48] 韩建平，贾宁凤. 土地利用与地形因子关系研究——以砖窑沟流域为例[J]. 中国生态农业学报, 2010, 18(5):1071-1075.
Han J P, Jia N F. Relationship between topographic factor and land use - a case study of Zhuanyaogou watershed[J]. Chinese Journal of Eco-Agriculture, 2010, 18(5):1071-1075.
[49] 哈凯，丁庆龙，门明新,等. 山地丘陵区土地利用分布及其与地形因子关系——以河北省怀来县为例[J]. 地理研究，2015, 34(5):909-921.
Ha K, Ding Q L, Men M X, et al. Spatial distribution of land use and its relationship with terrain factors in hilly area[J]. Geographical Research, 2015, 34(5):909-921.
[50] 武爱彬，刘欣，赵艳霞. 地形因子对浅山丘陵区土地利用类型分布与转换的影响——以燕山浅山丘陵区为例[J]. 农业现代化研究, 2014, 35(1):103-107.
Wu A B, Liu X, Zhao Y X. Influences of topographic on distribution and change of land use types in hilly region-taking Yanshan hilly region as an example[J]. Research of Agricultural Modernization, 2014, 35(1):103-107.
[51] 张丽萍，王小云，张赫斯. 沙盖黄土丘陵坡地土壤理化特性随地形变化规律研究[J]. 地理科学, 2011,(2):178-183.
Zhang L P, Wang X Y, Zhang H S. Evolution of physical and chemical characteristics of loess with different landforms in slope field under sand cover[J]. Scientia Geographica Sinica, 2011, (2):178-183.
[52] 徐国策，李占斌，李鹏,等. 丹江鹦鹉沟小流域土壤侵蚀和养分损失定量分析[J]. 农业工程学报, 2013, 29(10):160-167.
Xu G C, Li Z B, Li P, et al. Quantitative analysis of soil erosion and nutrient loss in Yingwugou watershed of the Dan River[J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(10):160-167.
[53] 吴昊. 秦岭山地松栎混交林土壤养分空间变异及其与地形因子的关系[J]. 自然资源学报,2015,(05):858-869.
Wu H. The Relationship between terrain factors and spatial variability of soil nutrients for Pine-Oak mixed forest in Qinling Mountains[J]. Journal of Natural Resources, 2015,(05):858-869.
[54] 王丽，王力，王全九. 不同坡度坡耕地土壤氮磷的流失与迁移过程[J]. 水土保持学报, 2015, 29(2):69-75.
Wang L, Wang L, Wang Q Q. The processes of nitrogen and phosphorus loss and migration in slope cropland under differene slopes[J].Journal of Soil and Water Conservation, 2015, 29(2):69-75.
[55] 郑子成，秦凤，李廷轩. 不同坡度下紫色土地表微地形变化及其对土壤侵蚀的影响[J]. 农业工程学报，2015，31(8):168-175.
Zheng Z Z, Qin F, Li T X. Changes in soil surface microrelief of purple soil under different slope gradients and its effects on soil erosion[J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(8):168-175.
[56] 张彩霞,杨勤科,李锐. 基于 DEM 的地形湿度指数及其应用研究进展[J]. 地理科学进展,2005,(06):116-123.
Zhang C X, Yang Q K, Li R. Advancement in Topographic Wetness Index and Its Application[J]. Progress in Geography, 2005, 24(6):116-123.
[57] Lewis G L, Holden N M. The Modification of Soil Moisture Deficit Calculation Using Topographic Wetness Index to Account For the Effect of Slope and Landscape Position[C]// 2012 Dallas, Texas, July 29 - August 1, 2012. 2012.
[58] Maduako I N, Ndukwu R I, Ifeanyichukwu C, et al. Multi-Index Soil Moisture Estimation from Satellite Earth Observations: Comparative Evaluation of the Topographic Wetness Index (TWI), the Temperature Vegetation Dryness Index (TVDI) and the Improved TVDI (iTVDI)[J]. Journal of the Indian Society of Remote Sensing, 2017,45(4):631-642.

[59] Zhu H, Hu W, Bi R, et al. Scale- and location-specific relationships between soil available micronutrients and environmental factors in the Fen River basin on the Chinese Loess Plateau[J]. Catena, 2016, 147:764-772.

[60] Zhu H, Zhao Y, Nan F, et al. Relative influence of soil chemistry and topography on soil available micronutrients by structural equation modeling[J]. Journal of Soil Science & Plant Nutrition, 2016, 16(4):1038-1051.